Data Transformation Is the CEO's Business

A massive, multiyear initiative to transform corporate data at Caterpillar demonstrates what top executives must do to unlock the power of their data assets.

By Barbara Wixom, Ogi Redzic, Brandon Hootman, Joaquin Rodriguez, Gabriele Piccoli, and Cynthia Beath

CATERPILLAR'S CEO HAD A PROBlem. Jim Umpleby had stepped into the chief executive role in 2017 with a vision of achieving more profitable growth by selling more services and parts to the company's heavy-equipment customers. Because offering real-time fleet management information services and selling parts online would depend on digital technologies, he set up a new division called Cat Digital. But the division's head soon had unwelcome news for Umpleby: The company didn't know its customers well enough to deliver on its goal. Customer data was siloed, fragmented, and, in the case of secondhand equipment, often entirely lacking.

That problem is one shared by countless leaders who see how digital can enable a growth strategy but are stymied by a legacy of fragmented, incomplete, and inconsistent data assets. They are frequently told that getting their data in order is a prerequisite for taking full advantage of new tools like artificial intelligence but too often see data transformation as an IT modernization project. They delegate the work to IT leaders and evaluate success based on cost, speed, and tool adoption. But when data is treated as IT infrastructure rather than as an enterprise asset, its impact is predictably limited. Companies that gain real value from their data actively involve the top management team in data transformation.

At Caterpillar, Umpleby and his team did more

Aeriform / Ikon Images

than simply sign off on the proposal by Ogi Redzic, chief digital officer of Cat Digital, to build a new enterprise digital data platform. In 2019, they demonstrated long-term commitment by giving Redzic's team a generous three years to build the platform, dubbed Helios, that would allow the company, its customers, and its dealers to see consistent and complete fleet information across all applications and interactions. Executives went on to redesign data governance, elevating data ownership to senior leaders. By 2025, the Helios platform was supporting e-commerce, fleet management, and predictive and preventive maintenance — and Caterpillar had grown its services revenue from $14 billion in 2016 to $24 billion in 2024.

What Really Drives Data Transformation

Caterpillar's experience underscores a critical lesson: Data transformation is not a purely technical exercise. Top company leaders must set a goal for the transformation in terms of business outcomes; give executives responsibility for data; commit resources to building an enterprise data platform; give all stakeholders a voice in the transformation; and direct strategic investments that take advantage of new data capabilities. Here, we'll explore each of these areas and the lessons learned from Caterpillar's experience.

1. Set (and monitor) an aggressive target for new revenue from the company's use of data. One particularly consequential action that Umpleby took helped to focus data transformation efforts early on: In 2019, he set the bold goal of reaching $28 billion in services revenues by 2026. That narrowed the company's attention on a small set of priority data initiatives and on closely tracking how data transformation led to increased services revenues.

Leaders initially prioritized developing a new tool, VisionLink, to help customers manage their fleets, and enabling customers to order parts via new e-commerce capabilities. Setting these priorities helped the team that was building the Helios platform focus on cleaning up the customer, customer contact, and equipment data needed for those solutions. Those three data types were known as "the trifecta" because they were key to answering important questions like "Which contact working for which customer should receive a sales or service offer?" and "Which contact working for which customer is responsible for replacing which machines?"

Top leaders regularly monitored Cat Digital's progress on those priority initiatives using three measures of value. Value enablement was measured by the number of accurate trifecta records on the platform. Value created for customers and dealers was assessed based on the number of VisionLink users, how often the application was used, and how quickly that usage was growing. Value realization was tracked as revenue attributable to the new solutions, such as parts sales. This value-based reporting approach helped business leaders connect progress made in data initiatives to impact on the corporate income statement.

By establishing clear business goals, leaders keep decision-making aligned with strategic objectives. Because foundational work can take years, tracking and reporting progress that has been made is important to sustain stakeholders' patience.

2. Give senior business executives ownership of data. Data becomes fragmented and inconsistent in many organizations because no one is accountable for its life cycle, quality, or reusability. At Caterpillar, establishing data as a corporate asset for which senior leaders were accountable focused attention on data quality. It also tied executives closely to the employees responsible for maintaining data products and related solutions.

To implement the new data-ownership policy, Cat Digital and the corporate IT team identified 14 enterprise data domains (such as engineering, finance, and human resources) and recruited about a dozen vice presidents across the company to own them. Some leaders represented functions such as finance or procurement, while others represented business units that had a strong stake in a particular data domain. Redzic took on customer, contact, and equipment data because those domains were key to services revenue. The vice presidents received information

THE RESEARCH

- The MIT Sloan Center for Information Systems Research conducted case study research at Caterpillar, a member of the MIT CISR research consortium since 2007.
- From July 2023 to December 2024, the authors conducted 56 interviews with 42 stakeholders. Interview participants also reviewed the case narrative as it was developed by the researchers and provided supplementary information.
- The authors supplemented the interviews with information from documents provided by Caterpillar and publicly available sources.

Before and After Caterpillar's Data Transformation

Caterpillar's data overhaul involved innumerable changes to how data is handled and used at the company. The following are three of the most consequential.

Before: Data was managed system by system.
After: Data is productized and managed as components on the central Helios platform. Each data product has a product owner who is accountable for its reliability and reusability. For example, a data product owner manages the data set that generates customer fleet lists and is responsible for its performance, wherever it is deployed. When a new solution team needs customer fleet lists, they don't rebuild data from scratch — they reuse the Helios data component. Over time, the inventory of digital components will grow, speeding assembly of new digital solutions.

Before: Data was cleaned via manual, one-off fixes specific to local business requirements.
After: A data science team builds data quality services using a variety of algorithms to detect and correct data quality problems. New, reusable services have increasingly automated data quality oversight processes. For example, one data quality service validates the serial numbers of equipment assets. Previously, when Caterpillar collected asset information, the serial number field often contained errors; it was not unusual for people to mistype the numbers, which could be hard to see or may have worn off. Because repair services and equipment sales require serial numbers, the data science team created an API-enabled asset validation service fueled by five machine learning algorithms for which Caterpillar eventually received a patent.

Before: Data was ingested despite containing anomalies and stored for use in applications. If applications or users detected data errors, the data may or may not have been corrected, via a process that was mostly manual.
After: A data operations team monitors a host of automated services that prepare data for use and resolve data problems as they arise. In the first stage of the process, data coming in from a dealer system or other source is stored in a raw-data staging area, with little transformation, if any. In the second stage, the services convert data into standard formats, fill in missing fields, correct invalid information, and create data objects, which are used as building blocks for use cases. In the third stage, services "bottle up" a set of data objects into a data product that meets the needs of a specific user or system. In the end, Helios generates data products, such as the customer master list, which serves as a single source of the company's most important information, and the VisionLink fleet list, which contains the assets that customers own, operate, or rent.

monthly about the quality of their data and were kept informed about the applications that used their data.

Cat Digital also established the role of *data product owner* — individuals who tracked value enablement and value creation metrics and were expected to drive data reuse and make data investments pay off. (See "Before and After Caterpillar's Data Transformation.") On the tech side, data product owners made sure that widely needed data, such as customer master lists and fleet information, were designed as reusable components. Owners of customer data came from the business side, bringing valuable domain expertise to the work of defining quality standards.

Owners of digital solutions were expected to generate customer value and service revenue. For VisionLink, that meant driving customer retention, customer satisfaction, and aftermarket parts revenue. Then owners of digital solutions collaborated with the owners of relevant data on creating data products, such as the fleet list, which tracked all the equipment each customer owned, operated, or rented. By maintaining a fleet list data product, Caterpillar ensured that its customers saw the same fleet list data everywhere.

Having vice presidents accountable for data assets, and establishing roles dedicated to managing data products and data solutions, made it clear that data was no longer an IT issue but an enterprise priority. Over time, executives understood why they had to be involved and took on the accountability and responsi-

bility that came with data domain ownership. At the same time, teams of all kinds started paying more attention to data decisions, knowing that their actions were subject to vice presidents' oversight. Executives did not step on each other's toes: No executive would invest in setting up a siloed customer database when one of their peers had clear responsibility for customer data.

3. Commit resources to building an enterprise data platform. Caterpillar began with a fragmented data landscape resembling that of many big, old companies. Independent dealers operated disparate systems. Multiple divisions maintained their own applications and databases. About 200 different interfaces connected dealer and corporate systems. Customers relied on multiple Caterpillar applications and often saw inconsistent information about their fleet. Cleaning up and consolidating such a landscape is not for the faint of heart. Many organizations bend to time pressures and cut corners on architecture and then fall back, yet again, on developing application-specific solutions optimized for immediate needs, exacerbating the architectural sprawl.

Caterpillar's leaders approached the cleanup with a realistic understanding of the time and money it would take to build a platform that could serve up the company's most important information consistently, quickly, and cost effectively. Cat Digital's work building the Helios platform ultimately resulted in the shutdown of eight legacy data platforms, reducing data management complexity by a factor of 30, by its calculation.

Over the three years spent building Helios, Redzic set milestones and regularly presented leadership with evidence that the group was meeting them. That sustained leadership's confidence over a relatively long time horizon as legacy systems were sunset and business areas moved onto the platform.

By investing upfront in an architecture for reusable data components and providing time for teams to stabilize foundational components, Caterpillar established the capability to deploy data products and solutions quickly and efficiently. (See "Caterpillar Data Architecture," p. 60.) By 2023, Helios was serving as a single source of data for digital services to customers. When a customer managed its fleet using VisionLink and used Caterpillar's e-commerce system to order parts, current data about the fleet was reflected in both systems. When a customer received preventive maintenance recommendations, those recommendations were associated with the bill of materials data that informed the e-commerce system. In 2024, the company saw record use of VisionLink, with thousands of customers newly onboarded.

4. Give internal and external stakeholders a voice in the data transformation process. Building solutions centrally and then pushing them to partners and business units often leads to resistance, fragmented adoption, and solutions that don't fit how work actually gets done. Caterpillar's leaders avoided this outcome by establishing channels by which independent dealers and a variety of business units could engage with the company's data transformation efforts. The approach ultimately helped speed the adoption of its digital solutions.

Dealer engagement was particularly important. Redzic established the Dealer Digital Council, which brought together interested dealers representing Europe, Africa, and the Middle East; the Americas; and the Asia-Pacific region. The council held two-day meetings each quarter to review digital product road maps. The discussions provided Caterpillar's leaders with direct feedback from the field and helped ensure that new digital services aligned with dealers' needs and operational realities.

A second group, the Enterprise Dealer Integration Council, focused specifically on data integration across the dealer ecosystem. Led by Caterpillar's vice president of dealer and customer support, the council was responsible for establishing standards for the data exchanged between Caterpillar and its dealers. All Caterpillar groups that sent data to or received data from dealers were represented on the council.

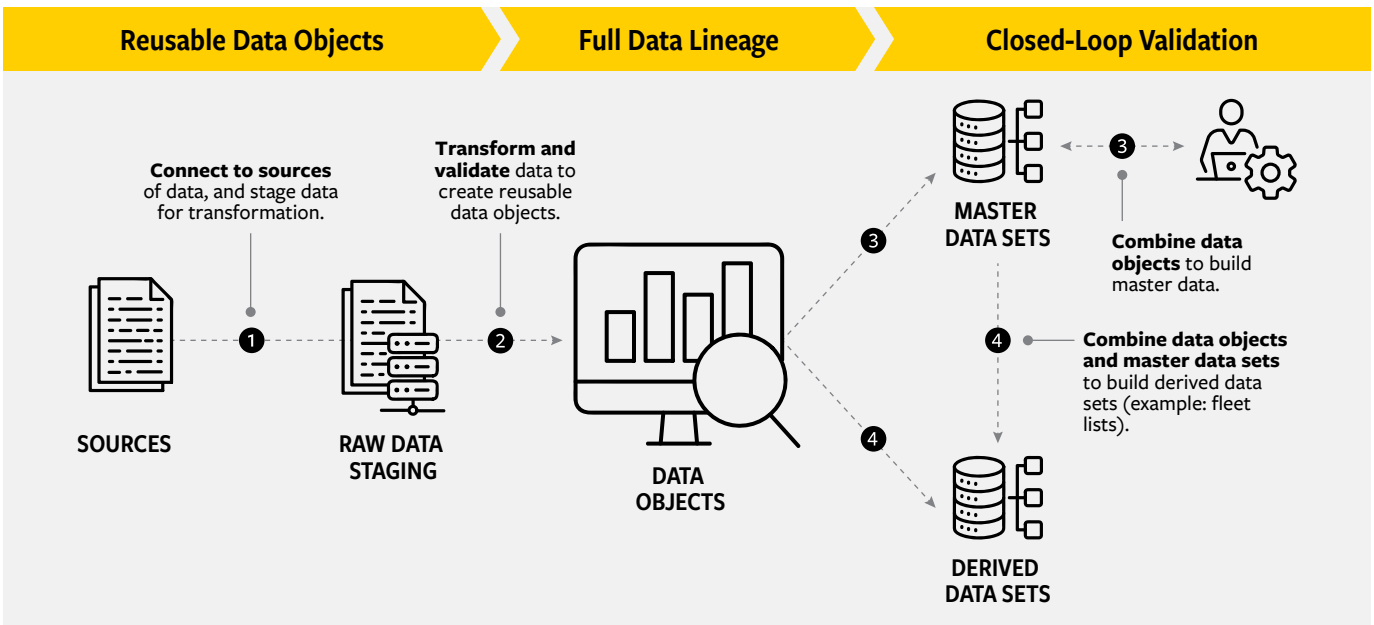
Internal business units were involved in shaping digital priorities via a new demand review board, where digital leaders met monthly with business sponsors proposing new initiatives. The board ranked project ideas based on their alignment with the Helios road map and their expected business benefits, cost, and required expertise.

The Demand Review Board created transparency about which initiatives would move forward and why. It also helped Caterpillar avoid a common pitfall of large enterprises: the reemergence of siloed applica-

Making VPs accountable for data assets made it clear that data was an enterprise priority.

Caterpillar Data Architecture

Caterpillar's Helios platform adheres to three key data architecture principles. The first is reusable data objects — data building blocks that can repeatedly be used to deliver consistent information to data consumers. The second is end-to-end visibility into data, from data source to data use. That means that at any point, the team can trace information back to raw data and the subsequent data objects and data quality processes that prepared it. The third is closed-loop validation — a key principle, given that the platform is self-reliant for data quality. The platform continuously monitors and analyzes data quality as data moves through the platform processes, and kicks out problems for review.



Source: Caterpillar

tions and redundant data pipelines. By reviewing initiatives centrally, the board nudged teams to harness Helios's reusable capabilities rather than build new, isolated solutions. The result was fewer applications overall and greater consistency in data, user experience, and platform utilization — ultimately reducing enterprise spending.

The value of the board was evident in the development of the Cat Vantage Rewards program (recently relaunched as Cat Rewards). Leaders of Cat Financial proposed expanding the company's rewards program to allow customers to earn points from online purchases made with a Cat credit card. Implementing the idea required that data from Caterpillar's e-commerce systems, dealer invoicing systems, and customer accounts be connected — a task that would have been extremely complex in the company's previous, fragmented data environment.

Instead of building a solution from scratch, Cat Financial took advantage of the Helios capabilities already in place. Using Helios customer master data and existing data objects related to invoices, the teams quickly developed services that identified eligible transactions and calculated reward points. Within months, the solution had progressed to testing and integration with Caterpillar's broader digital rewards experience. According to leaders within Cat Financial, implementing the program would likely have taken years under the company's earlier architecture.

Caterpillar's senior leaders learned the importance of ecosystem involvement and impact. By involving dealers, business units, and product teams in shaping the company's data and digital initiatives, they drove viable solutions and desirable outcomes. Data transformations succeed not only because of strong architecture but also because of strong relationships. Organizations that give their ecosystems voice dramatically increase the likelihood that new capabilities will be adopted and add value across the enterprise.

5. Build on new data capabilities with strategic investments such as AI. While working on the data transformation, Caterpillar began building its AI expertise by using machine learning tools to improve data quality. Its data science team worked closely with business domain experts on models that could identify and correct anomalies in incoming data, such as incorrectly entered equipment serial numbers or inconsistencies in customer records. Besides automating previously manual work, the models allowed the company to maintain high quality standards as the volume and complexity of data increased. Just as important, the development of the models educated business collaborators and senior leaders on possible use cases for AI.

Cat Digital also engineered new AI capabilities directly into Helios as generative AI emerged. The team created vector data stores to house large volumes of unstructured information, such as equipment manuals and service records, so that the AI systems could quickly search them. That allowed employees and customers to ask questions and receive precise answers based on information included in company documents. Cat Digital created prompt libraries for generating service recommendations and summarizing equipment performance. It also built agent orchestration services to coordinate multiple AI agents and digital services so that they could work together to complete tasks — for example, analyzing machine telemetry, identifying a maintenance issue, generating a service recommendation, and notifying a dealer or customer of a problem. To ensure that AI development remained closely tied to business objectives, the AI Digital Product Council evaluated potential use cases and prioritized those most likely to create value for customers, dealers, and the enterprise.

Key Points for CEOs Leading a Data Transformation

Caterpillar's experience makes a compelling case for why transforming data to unlock the value of those assets requires attention and leadership from the CEO suite. Executives who know that their organization's data requires similar transformation — most urgently, so they can take advantage of AI — should keep the following in mind as they hone their strategy and leadership approach.

- Set a stretch business goal that the data transformation will serve, and use that goal to focus people's efforts and keep the overall project on track.
- Tie data investments to revenue growth, margin expansion, risk reduction, or customer value creation rather than simply tracking infrastructure spending. Monitor a set of metrics that clarifies the causal chain from platform investment to business aspiration.
- Require that business leaders assume data ownership; don't assign this to technology leaders. Establish product management roles — for data assets and data solutions — to manage the deployment of reusable data and track data payoffs.
- Invest in a data platform, not another data silo. Monitor progress using enterprise KPIs like data reuse and data liquidity (that is, how available data is for immediate use). Give people the time they need to establish foundations rather than add to sprawl.
- Involve partners and business units in informing and shaping data transformation. Establish mechanisms to identify and proactively manage potential conflicts.
- Use AI to automate data management tasks while establishing the company's data foundations.

EVERY LARGE ORGANIZATION TODAY manages vast data assets, but few extract their full value. The difference lies in executive leadership. When CEOs talk about data in earnings calls, participate in governance discussions, and hold leaders accountable for data quality, the organization listens. In an era where competitive advantage increasingly depends on insight, integration, and intelligent automation, companies that treat data as chiefly the responsibility of the IT function will fall behind. Those that treat it as a corporate asset — and lead accordingly — will define the next decade of performance. ■

Barbara Wixom *is a principal research scientist at the MIT Center for Information Systems Research (CISR).* **Ogi Redzic** *is a senior vice president and chief digital officer at Caterpillar.* **Brandon Hootman** *is vice president of physical AI platforms and AI autonomy at Caterpillar.* **Joaquin Rodriguez** *is an assistant professor at the Grenoble Ecole de Management and a research fellow at CISR.* **Gabriele Piccoli** *is the Edward G. Schlieder Chair of Information Sciences and a member of the Cultural Computing group at the Center for Computation and Technology at Louisiana State University.* **Cynthia Beath** *is a professor emerita at the McCombs School of Business at the University of Texas at Austin.*

Reprint 67426. For ordering information, see page 4. Copyright © Massachusetts Institute of Technology, 2026. All rights reserved.

https://doi.org/10.63383/igDH4176